11
Subcellular Targeting and Purification of Recombinant Proteins in Plant Production Systems

MAURICE M. MOLONEY[1,2] AND LARRY A. HOLBROOK[1]

[1]*Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada;* [2] *Norman Borlaug Institute for Plant Science Research, De Montfort University, Leicester, UK*

Introduction

The expression of foreign proteins in plants has become a routine exercise for a wide range of dicotyledonous (Fraley *et al.*, 1986) and more recently monocotyledonous (Gordon-Kamm *et al.*, 1990; Vasil *et al.*, 1993) plants. The vast majority of such work has been focused on improvement of agronomic traits such as herbicide tolerance (Mazur and Falco, 1989) and disease or pest resistance (Staskawicz *et al.*, 1995; Koziel *et al.*, 1993), while other work has been devoted to altered crop quality in relation to fruit ripening (Smith *et al.*, 1988) or lipid profile (Töpfer *et al.*, 1995; Voelker *et al.*, 1992). In all these cases the novel trait is produced by the expression or suppression of a protein but normally there is no need, apart from analytical purposes, to separate and purify the novel protein from the plant.

The development of plant biotechnology has followed a different path from many other fields of biotechnology where the purpose of expressing a protein, for instance in micro-organisms or mammalian cells, has been for the production of a purified product. This has brought about several notable successes including erythropoietin (Goto *et al.*, 1988), human insulin and a wide range of natural and recombinant antibodies.

As it has long been possible to engineer plants to produce recombinant proteins which modify the plant's phenotype, it is reasonable to pose the question as to why plants have not been used widely for the production of recombinant proteins, for which they have a number of potential advantages:

1. Plants are in many cases inexpensive to produce, generating hundreds of kilograms of (their native) proteins per hectare
2. Plants are capable of storing proteins stably in a variety of specialized organs (e.g. seeds, tubers)

Biotechnology and Genetic Engineering Reviews – Vol. 14, April 1997
0264–8725/94/12/321–336 $20.00 + $0.00

3. A wide infrastructure exists for the processing of plant-derived products (e.g., crushing for seed protein and oil, wetmilling for proteins and glucans).

Despite these and other attractive features, it is interesting that plants have not been used as hosts for recombinant protein production on a commercial scale. The reasons for this include:

1. Concern about the environmental safety of plants producing distinctly non-plant products
2. A preference for production in sterile, stainless steel or glass containers
3. The potential costs of purification which may offset the attractive economics of plant cultivation.

The first two issues are very significant, but beyond the scope of this article. The question of environmental safety of transgenic plants is under broad scrutiny (Crawley *et al.*, 1993). The second consideration is a preference which is certainly not an absolute. Highly purified pharmaceuticals for human use have been and are produced from such diverse sources as Yew bark harvested in forests and even equine urine. The primary factors which promote using non-conventional sources of proteins or other products for use in health care are uniqueness and relative cost.

It is in this context that we will discuss the third factor influencing the use of plants as hosts for production of recombinant proteins, the cost of purification of the product. It is noteworthy that the plant-derived recombinant proteins most likely to be used commercially involve no purification. They utilize the property of a seed to act as a carrier of a protein which need not be purified. This is typified by the work of Pen *et al.* (1993b) who produced an *Aspergillus* phytase in seeds of tobacco. These seeds could be mixed directly in chicken feed rations with other non-recombinant seeds, with the result that phytate was hydrolysed to useable phosphate and inositol when the feed was consumed by the animal. This principle is probably more widely applicable and would render plants useable as hosts for recombinant proteins without imposing a purification step.

The broadest application of plants as hosts for recombinant protein production will eventually involve purification of the desired protein. Its separation may be greatly facilitated if the protein is sequestered in a particular cellular compartment. If the cellular compartment has a valuable feature (e.g., protein stability or ease of isolation) this could give rise to an inexpensive production system. There are several subcellular trafficking strategies which have proved to be useful in this regard. These include use of all aspects of the endoplasmic reticulum (ER)-secretion pathways, resulting in sequestration of proteins in ER, tonoplast or the apoplast.

Alternatively, other organelles are targeted, including chloroplasts and oil bodies. In some cases a new structure, for example a virion, may also serve as an easily separable entity for further protein purification, although this latter idea will not be discussed here.

Use of secretion pathways for subcellular targeting

The secretion pathway of animal, fungal and plant cells offers singular advantages for the production of recombinant proteins. The secretion pathway in plants regulates and

determines the passage of polypeptides to tonoplast-derived protein bodies, ER derived protein bodies or indeed, secretion into the apoplastic space. The advantages of using the secretion pathway are manifold. Firstly, many polypeptides of pharmaceutical interest are secreted proteins in their source organism. They may undergo specialized folding and post-translational modification that requires components of the ER such as BiP chaperones or glycosylation enzymes. Secondly, secretion into one of the cellular compartments mentioned above may also separate the desired protein from proteases likely to catalyze its breakdown. Thirdly, sequestration into certain cellular compartments may facilitate recovery of the protein on extraction, if that compartment can be preferentially enriched during processing.

Use of the plant secretion system, however, is not without difficulties. The most significant of these is the question of glycosylation. N-linked glycosylation in plants occurs initially in the ER. Subsequent modifications of simple glycans to complete or complex glycans occurs in the Golgi apparatus. In plants, the production of complex glycans involves the hydrolysis of mannose residues by glycosidases and the addition of plant-typical sugars such as xylose or fucose (Faye, 1988).

Studies on the complex glycans which are frequently added to glycosylated plant proteins indicate that they are usually quite immunogenic in animals. Faye *et al.* (1993) actually used xylose-specific recognition as the basis of an affinity purification scheme for recovery of antibodies against xylose containing epitopes.

These properties would appear to obviate using the secretory pathway in plants for production of proteins for human therapeutic use. However, von Schaewen *et al.* (1993) screened an M2 mutant population of *Arabidopsis thaliana* using antisera against complex glycans. They isolated a mutant which was blocked in the accumulation of proteins containing complex glycans. Cellular extracts were found to lack N-acetyl glucosaminyl transferase 1, the first enzyme required in the pathway to synthesis of complex glycans. These workers crossed to mutant plant with a transgenic *Arabidopsis* expressing the glycoprotein phytohaemagglutinin (PHA) and obtained a novel plant producing PHA without any complex glycan, but rather a mannose-rich glycosylation pattern. Interestingly, there was no apparent phenotype to this mutation, suggesting that it may be possible to produce a special plant host lacking complex glycosylation as a production vehicle for proteins of therapeutic interest.

TARGETING TO PROTEIN BODIES

The possibility that recombinant proteins could be targeted to protein bodies was first raised by Sengupta-Gopalan *et al.* (1985). These workers showed that a phaseolin genomic clone comprising promoter, coding region including introns and about 1200 bp 3′ of the coding sequence could be expressed in tobacco after *Agrobacterium*-mediated transformation. The gene expressed in a strongly seed-specific manner and yielded a polypeptide which underwent glycosylation and targeting to protein bodies. Indeed, expression of a single copy insert of the phaseolin gene in tobacco resulted in about 3% of total seed protein being phaseolin. This work was the first to suggest that heterologous polypeptides in plants might be correctly processed through the secretion system and reach their 'correct' subcellular compartment.

Hoffman *et al.* (1988) attempted to extend this idea to include genetically modified storage protein genes. They introduced into tobacco a β-phaseolin gene into which

was inserted a 45 bp nucleotide fragment encoding a methionine-rich stretch derived originally from a zein (Pederson *et al.*, 1986). This stretch was incorporated near the 3′ end of exon 3 of the phaseolin genomic clone in a region with predicted α-helical structure. Expression of this 'himet' phaseolin was very similar to unmodified phaseolin at the mRNA level. However, tobacco seeds were incapable of accumulating the himet protein, probably due to protein instability and turnover.

This result, although apparently discouraging, rather reflects the precision with which modifications must be made in order to produce recombinant protein fusions which can accumulate. A key study in this respect was that of Vandekerckhove *et al.* (1989). These workers investigated the property of the 2S albumins from the tribe *Brassicae* as possible carriers of recombinant proteins. Because of findings by Krebbers *et al.* (1988), who showed that the 2S albumins from *Arabidopsis* and *B.napus* were all cleaved near their C-termini as part of the processing of large and small subunits, Vandekerckhove *et al.* recognized that simplistic modifications such as tagging the recombinant protein to the C-terminal end of these albumins was unlikely to work. One region in the 2S albumins was, however, shown to be variable both in length and sequence. This was found between the 6th and 7th cysteine in 2S albumins from *Arabidopsis*, *Brassica* and the distantly-related *Bertholletia excelsa* (Brazil nut) (Ampe *et al.*, 1986). Thus, Vandekerckhove *et al.* (1989) produced a gene construct encoding an internal protein fusion of an *Arabidopsis* 2S albumin and the pentapeptide *Leu*-enkephalin between these cysteines. This gene fusion was introduced into *Arabidopsis* and *B. napus* and expressed correctly as a fusion. The recombinant protein was targeted and accumulated in protein bodies. These workers subsequently showed that the peptide could be released from the mature recombinant 2S albumin by tryptic cleavage and the extra lysine could be removed using carboxypeptidase. Yields of up to 206 nmol of peptide per gram of seed were found.

Promising as these results may appear, there are some significant limitations. The cleavage of a protein enclosed with flanking sequence generally results in an extra amino acid at the C-terminus of the peptide. The removal of this on a large scale could be costly and renders the process somewhat complicated. Perhaps more severe is the problem of size limitation. Although no published reports have determined the precise limit, the natural range of sizes of this presumed loop in the 2S albumins probably defines the limits of permissible modifications. This would suggest that a maximum size of about 40 amino acids would be tolerated. Obviously, it is hard to predict whether a particular sequence would be tolerated. This could only be determined empirically at this time.

Despite some of these limitations, the potential use of storage proteins as 'carriers' of recombinant proteins still merits further study. Of particular interest for further work in this direction are the zeins and other prolamines. It has been known since the work of Hoffman *et al.* (1987) that the 15 KDa zein could be expressed in heterologous seeds such as tobacco and that the protein would accumulate in protein bodies. This finding, on its own, could be interpreted in the same light as phaseolin expression in tobacco seeds (Sengupta-Gopalan *et al.*, 1985). However, more recent studies by Bagga *et al.* (1995) suggest that the storage of zein is not confined to seeds and clearly undergoes targeting and processing distinct from a 7S globulin such as phaseolin. In their experiments, Bagga *et al.* (1995) showed that when a zein was expressed constitutively using a 35S CaMV promoter, the protein accumulated in leaves in well-

defined protein bodies which appeared to arise from RER.

This was distinct from a 35S-promoter-β-phaseolin gene which permitted seed-specific accumulation, but not accumulation in leaves. It was suggested that in leaves the β-phaseolin is targeted (correctly) to vacuoles, but that it undergoes rapid turnover in the proteolytic environment of the vegetative tonoplast. The observation of zein accumulation in leaves in well-defined protein bodies clearly suggests that zeins could be used to increase protein content in forage crops, but also raises the question of expressing recombinant zeins as fusions with other desired protein. Zeins undergo far less subcellular processing than do 2S albumins, and also exist in a variety of molecular weights. It may be possible to introduce C-terminal tags into zeins without post-translational loss, although this has never been tested. Given the rather precise requirements for zein solubilization, purification from seeds or leaves of a recombinant form may prove to be easy to perform, even on a large scale.

SECRETION INTO THE APOPLAST

Proteins which possess a signal sequence specifying co-translation and ER association frequently undergo secretion into the apoplastic space. In the absence of other traffickers or retention signals, the 'default' pathway for subcellular targeting is directed to the extracellular space. In addition, signal sequences such as that derived from the tobacco pathogenesis-related protein, PR1a, if used in conjunction with heterologous proteins, can direct these proteins into the extracellular space (Pen *et al.*, 1993a).

In studies on the production of recombinant proteins in plants, secretion of proteins into the apoplast may be very advantageous. The advantages include: sequestration of the desired protein, ER and Golgi processing of the polypeptide, and exposure of the nascent polypeptide to BiP proteins, the natural chaperones associated with ER. Finally, deposition of the protein in the apoplast could act to protect it from enzymatic degradation.

There are numerous examples of successful use of such secretion approaches. In some cases, exploitation of the secretory pathway will be necessary either for correct folding of the desired protein or for glycosylation. As discussed above, glycosylation patterns in plants may be quite distinct from those found in mammals, particularly with fucose and xylose residues which may be linked to GlcNAc or core mannose (Faye *et al.*, 1989). However, the result of Von Schaewen *et al.* (1993) suggests that undesirable glycosylation patterns could be avoided in transgenic plants producing proteins intended for therapeutic use, either through the use of a mutant host, or by an alternative form of target gene suppression such as antisense RNA or co-suppression.

A large number of recombinant proteins have been tested in such secretion systems. One of the earliest examples was the production in plants of antibodies designated 'phytoantibodies' (Hiatt *et al.*, 1989). While interesting for their individual applications, studies on phytoantibody production also illustrate a number of general points about secreted recombinant proteins.

Phytoantibodies

Hiatt *et al.* (1989) produced transgenic tobacco plants expressing cDNAs encoding in plants γ or κ chains of a mouse monoclonal catalytic antibody (6D4). Separate

transformants expressing either heavy or light chain cDNAs were crossed sexually to yield plants expressing both chains. This resulted in plants capable of producing properly assembled antibodies. Successful accumulation of the γ or κ chains and assembly was dependent on a functional secretion leader sequence. Interestingly, that sequence used in Hiatt *et al.*'s (1989) experiment was the natural mouse secretion leader, indicating that plants correctly recognize and use these signals. Düring *et al.* (1990) performed similar experiments using cDNAs encoding light and heavy chains of the murine immunoglobulin BI-8. They substituted secretion leaders of the B-I-8 polypeptides with one derived from a barley α-amylase gene, and they too found assembled antibody in plant cells, although immunogold labelling suggested that most of the phytoantibody was located in the ER, *not* in the apoplast. Surprisingly, a significant amount of assembled antibody was also found in the chloroplast. Secretion of antibodies in plants may, however, lead to the occurrence of some unique changes. For example, in the case of the 6D4 murine antibody, glycosylation occurred. This glycosylation was probably complex and plant-specific (see above) as judged by endoH digestions and lectin binding assays (Hein *et al.*, 1991). However, this did not alter the kinetic parameters (K_m, K_i, V_{max} and K_{cat}) of the phytoantibody.

With the advent of single-chain antibody binding proteins (scF_v; (Bird *et al.*, 1988)) and the results of Hiatt *et al.* (1989) and Düring *et al.* (1990), it was a natural extrapolation to test the expression and functionality of single-chain antibodies in plants. Owen *et al.* (1992) produced a scF_v of an anti-phytochrome monoclonal (light-linker-heavy) and expressed this in transgenic tobacco plants. Expression of this recombinant scF_v occurred in several plants. Protein accumulation in these plants was not high, although it could be detected immunologically and phenotypically by a significant change in red-light sensitivity of germination caused presumably by interference with endogenous phytochrome function. Firek *et al.* (1993) demonstrated that much higher levels of the anti-phytochrome scF_v could accumulate in the apoplast if a suitable secretion leader is used.

Overall, it may be concluded that for the *production* of phytoantibodies in large quantities, targeting through the secretion pathway is required. This may be because of a need for disulfide isomerase and the less reducing environment of the ER, the presence of molecular chaperones such as BiPs in plants (Li *et al.*, 1993) which catalyze correct folding or even the segregation of the foreign protein outside the cell away from cytoplasmic or vacuolar protease. The remarkable versatility of plant systems in the correct assembly of immunoglobulins has recently been highlighted by Ma *et al.* (1994; 1995). These workers, using a monoclonal antibody (Guy's 13) against a surface epitope of *Streptococcus mutans* produced not only correctly assembled Guy's 13 in transgenic plants, but when a modified gene specifying a hybrid IgG/IgA heavy chain was produced, this also underwent correct assembly. The objective was to produce a fully assembled secretory immunoglobulin A (SigA) or close approximation by domain swapping. Normally this comprises four polypeptides: the γ or κ chains, a small joining peptide (j) and a polypeptide secretory component. In mammals, this assembly requires two cell types (plasma and epithelial). Surprisingly, transgenic tobacco plants expressing the four required polypeptides with appropriate secretion leaders were capable of assembly of the hybrid SigA-G antibody into a functional molecule which properly recognised the *Streptococcal* antigen. This

achievement argues that plants could be used favourably for production of a variety of antibodies for therapeutic or diagnostic use.

Secreted industrial enzymes

Although production of phytoantibodies might represent high unit value recombinant proteins, there are now several examples of enzymes for industrial and food use which may be produced economically in plants by exploiting secretory pathways. For example, Pen *et al.* (1992) demonstrated that expression of a *Bacillus licheniforms* alpha-amylase in transgenic tobacco seeds provided a functional enzyme at about 0.3% of soluble protein. Interestingly, passage through the secretory system resulted in glycosylation, giving a protein of 64KDa rather than 55.2KDa as produced in the bacterium. Despite this significant change, there was no alteration in the activity or specificity of the amylase produced. In this study, two different secretion leaders were tested, one of plant origin and the other the native bacterial secretion sequence of α-amylase. Surprisingly, the bacterial secretion leader functioned well in the eukaryotic system. A second example of a valuable industrial enzyme produced in plants via the secretion system is phytase. Phytase is an enzyme which catalyses the hydrolysis of phytate to myo-inositol and inorganic phosphate. The phosphate in phytate is inaccessible to monogastric animals and phosphate must normally be added to their feed. This problem is further compounded by the fact that these animals then excrete phosphate in phytate which results in eutrophication of surface waters.

Phytase may be obtained form a variety of fungi and bacteria. Pen *et al.* (1993b) expressed a gene encoding an *Aspergillus niger* phytase in tobacco plants. This yielded seeds containing a fully active phytase in the extracellular space. The fungal phytase underwent glycosylation. While this normally occurs in *Aspergillus*, the glycosylation in plants was distinct, involving complex glycosylation discussed earlier. Once again, however, this did not affect the activity. This example is particularly attractive as an economically viable model since the enzyme was active *in the seed* and could be added without purification to animal feed ration. In this 'formulation', the phytase had the desired effect and obviated the need for further added phosphate to the feed.

OTHER PROTEINS FOR THERAPEUTIC OR BIOLOGICAL APPLICATIONS

In addition to the production of phytoantibodies, plants have also been exploited to produce specific antigens which could be used for oral immunization. A recent study by Haq *et al.* (1995) sought to produce the 11.6 KDa β (binding) subunit from the *Escherichia coli* enterotoxin in plants. Normally, this β subunit called LT-β forms a stable cyclic pentamer which binds to the G_{M1} ganglioside of mammalian epithelia. Antibodies against LT-β can interfere with this binding; thus, in principle, the protein may be used as a vaccine for immunization. Haq *et al.* (1995) expressed this LT-β, including an ER retention signal (SEKDEL) at the end of the specified polypeptide. On transforming tobacco and potato plants, both were found to express the polypeptide and this underwent oligomerization to produce an assembled protein capable of binding the natural ligand, the G_{M1} ganglioside. Accumulation of the LT-β polypeptide was higher when the ER retention signal was present than in controls lacking the

SEKDEL motif. Not only did the plants produce an authentic oligomeric protein, but this protein was found to function well as an antigen. In fact, mice orally inoculated with leaf extracts, or indeed, mice allowed to feed on potato tubers expressing the antigen produced neutralizing antibodies against the antigen. These experiments illustrate that the potential for using plants as vehicles for vaccine production, particularly for mucosal immunization, is high and it is expected that wider use may now be made of this effective strategy.

The secretion pathway of plants has been widely used for a variety of potentially bio-active molecules. The first example was the production of Human Serum Albumin (HSA) (Sijmons *et al.,* 1990) in transformed potato leaves. The HSA produced was secreted into the apoplast even under the influence of the native mammalian leader. Through a series of chromatographic and electrophoretic tests it was determined that the HSA was chemically similar to natural HSA, although no evaluation of folding or biological activity was attempted.

Another example of a protein with potential therapeutic value is erythropoeitin (Epo). Epo is a natural glycosylated cytokine involved in the regulation and maintenance of steady-state levels of erythrocytes in blood. Mature Epo has three N-linked glycans at positions 24, 38 and 83 (Krantz, 1991). The glycans are not required *in vitro* for activity, but are involved in half-life *in vivo*. Matsumoto *et al.* (1995) expressed human Epo in tobacco suspension cultures. The native Epo cDNA was used and thus secretion depended on the mammalian leader sequence.

Chloroplast targeting

While the secretion pathway has been a major focus of work to produce recombinant proteins in plants, there has also been an interest in exploiting the unique properties of plastids. Alteration of the genetics and biochemical functions of chloroplasts could prove critical in a number applications. Specifically, the chloroplast is the site not only of photosynthetic reactions, but also of lipid biosynthesis, and of starch and amino acid metabolism. Nuclear encoded genes specifying protein products which are functional in chloroplasts include a domain encoding transit peptide sequences (Keegstra, 1989). These can be used to direct entry of heterologous polypeptides into the plastid. In consequence, efforts to alter lipid, starch or aminoacid biosynthesis can, therefore, be approached by nuclear transformation (Töpfer *et al.,* 1995). Two recent examples of effecting plastid metabolism for unique biosyntheses after plant nuclear transformation are the production of high levels of free lysine in canola and soybean seeds (Falco *et al.,* 1995) and the synthesis of the biodegradable plastic, polyhydroxybutyrate (PHB), in *Arabidopsis thaliana* (Nawrath *et al.,* 1994).

An alternative approach is to transform directly the plastid genome (plastome). This has been achieved with vectors harbouring prokaryotic sequences and selectable markers. A recent example in this field is demonstrated by Zoubenko *et al.* (1994), in the development of pPRV or Plasmid Repeat Vectors. These have flanking plastid homologous sequences for homologous recombination, multiple cloning sites, spectinomycin selection and no readthrough transcription from outside promoters when in the correct orientation. The production of 'transplastomes' thus follows biolistic delivery of these vectors and appropriate tissue culture growth and selection. Since a cell can have 10–50, 000 plastome genomes, expression can be very substantial.

A frequent objective of agricultural biotechnology has been to improve the amounts of amino acids which are deficient in seed meal of animal feeds. Lysine, as an essential amino acid, is relatively sparsely represented in prolamins such as zeins, the key storage proteins in corn. The usual approach has been to increase lysine by genetically engineering heterologous high-lysine peptide fragments into the crop seed storage proteins. This can however, as discussed earlier, impact negatively upon proper folding and protein body formation.

Falco *et al.* (1995) took a unique approach to increasing free lysine by altering the feedback control of the aspartate-family biosynthetic pathway in the chloroplasts of both canola and soybean. Lysine is derived from aspartate, and exerts feedback regulation on aspartokinase (AK) and dihydrodipicolinic acid synthase (DHDPS), two of the enzymes in the biosynthetic pathway. Falco and colleagues exploited lysine-insensitive mutant genes for these two enzymes, derived from *Corynebacterium* and *E. coli*, transcriptionally regulated by the β-phaseolin promoter and targeted using a chloroplast transit peptide (soybean small subunit RuBP carboxylase). The resultant transgenic plants had variable but increased levels of free lysine in their seeds. Since this is free lysine, there were also increased intermediates of lysine catabolism (α-amino adipic acid in canola, and saccharopine in soybean). Extensive testing for seed viability, germination and for animal nutritive value still have to be determined. High lysine corn transformants, which would be the major crop target for this technology, have also been achieved. Therefore, the possibility of affecting an endogenous pathway in plastids appears to be very feasible by using engineered nuclear genes with transit peptides. Falco *et al.* also estimate, based on lysine supplement costs and their yields of free lysine, that they have potentially increased the soybean meal value by 30%–35%.

One non-protein/aminoacid product which may be made in plants is worthy of discussion in this context. This is the production of poly hydroxy alkanoates in plants. Although apparently straying from the original topic of this review (recombinant proteins), mention is relevant because of the potential value of the product and the need to investigate subcellular targeting of the biosynthetic enzymes.

Poly hydroxy alkanoates are storage polymers found in a variety of bacteria, including *Alcaligenes* and *Rhodococcus*. These biodegradable polyesters require three enzymes for their biosynthesis from acetyl CoA: 3-ketothiolase, acetoacetyl-CoA reductase and PHA synthase. In bacteria, polymerization to various chain lengths results in the formation of highly lipophilic PHA granules.

In initial experiments, Poirier *et al.* (1992) expressed genes for PHA biosynthesis without transit peptide sequences. This allowed some PHB to accumulate in the plants as granules. However, the granules were in a range of cellular compartments including nuclei, the yield was low and the plants were stunted. A more recent report (Nawrath *et al.*, 1994) revealed that a dramatic improvement in yield and plant viability could be derived by targeting all three required gene products to the plastid by using the pea chloroplast transit peptide from RuBP carboxylase small sub-unit. This resulted in a 100-fold increase in yield from the original experiments, generating up to 14% of the leaf dry weight as poly hydroxy alkanoates. The viability of the plants and their seed production also suggested that the plastid acetyl-CoA pool was effectively regulated even with the high metabolic demand for a foreign biosynthetic process. This may prove to have economic advantages in bioprocessing and production of environmentally beneficial materials that could be sequestered in this cellular organelle.

Therefore, the plastid is an interesting target for manipulation of biosynthetic processes. Both of the examples given above utilized bacterial genes having plant promoters (either constitutive or seed-specific) and transit peptide sequences for nuclear encoded gene products. In one case, an endogenous pathway of at least two mutant genes was altered, and in the other, a totally new pathway that involved three genes was incorporated. Thus, not only can normal physiological functions be re-routed, but novel biosynthetic pathways may be integrated into the existing pathways of the organelle.

Synthesis of recombinant proteins on lipid bodies in seeds

Most of the successful work on the synthesis and subcellular compartmentation of recombinant proteins in plants has focused on protein body or apoplast targeting. These systems have undeniable merits, which have already been discussed, but they require some ingenuity in order to exploit them commercially as the method of gene expression does not solve the question of protein recovery and ease of purification.

The present authors have taken a somewhat different approach by investigating the properties of oil bodies rather than protein bodies. In all seeds which accumulate even small amounts of triglyceride or other storage lipid, the lipid is sequestered in oil bodies. These oil bodies are variously named spherosomes, lipid bodies, oil bodies or oleosomes. Oil bodies are not merely oil globules. They are distinct organelles comprising storage lipid and a half unit membrane (Yatsu and Jacks, 1972). In seeds (Huang, 1992) and pollen grains (Roberts *et al.*, 1993), but not fruit (Ross *et al.*, 1993) they are also embedded with a high concentration of proteins named oleosins. These are unusual proteins containing in their structure the longest stretch of highly lipophilic residues in proteins studied to date. This lipophilic stretch assures that once associated with oil bodies the oleosins remain tightly bound because of the preferential interactions of the lipophilic residues with the storage lipid in the oil body. A variety of models (Huang, 1992; Murphy, 1993) have been proposed for the configuration of the association of oleosins with oil bodies. These models will not be further discussed here as the issue remains unresolved. However, it is clear that the central lipophilic core of the oleosin protein is normally situated inside the oil body, whereas the more hydrophilic or amphipathic N- and C-termini are on the cytoplasmic side of the oil body membrane. This has been established *in vitro* using proteins digestion/protection experiments (Hills *et al.*, 1993) and *in vivo* by tagging oleosins with heterologous polypeptides (Van Rooijen and Moloney, 1995). Oleosins accumulate in most oil seeds at levels that are comparable to storage proteins. They are certainly commonly found comprising 5%–10% of total seed protein. Huang (1992) has calculated from such estimates the likely density of oleosins on oil bodies. Depending on their configuration on the surface oleosins might easily form a 'shell' of protein on the oil bodies. This idea is supported by experiments (Huang, 1992) in which phospholipase was shown to be capable of disrupting oil bodies *in vitro* only after a trypsin pretreatment, which presumably degrades oleosins on the surface. This 'shell' of tightly associated proteins on oil bodies probably explains why oil bodies remain intact (but highly compressed) during seed desiccation rather than coalescing, as would simple oil droplets.

The presence of oleosins also seems to make oil bodies fairly robust so that when

seeds are milled in aqueous buffer the oil is released as a scum of *intact oil bodies,* which generally rise to the surface after standing, centrifugation or through manipulation of buffer pH. Oil bodies are therefore easily separated from the other cellular contents on a small or indeed larger scale (Jacks *et al.,* 1990). Electrophoretic comparison of total seed extracts of oil bodies derived from the same seed show a dramatic degree of purification simply through flotation centrifugation.

It was in this context that the present authors reasoned that oil body proteins might serve as useful carriers for desired proteins and peptides. The targeting of oleosins to oil bodies is certainly less well studied than targeting of storage proteins. Nevertheless, it is clear that oleosins are co-translationally processed on ER and delivered to oil bodies by a mechanism which has not yet been elucidated. Studies on the comparative molecular biology and biochemistry of oleosins from diverse species have suggested that the lipophilic central core is quite conserved throughout evolution. On the other hand, the N- and C-termini seem to have diverged at the amino-acid sequence level. This has raised the question of the relative roles of the N- and C-termini in oleosin targeting and function. We have recently shown (Van Rooijen and Moloney, 1995b) that the C-terminus may be unimportant for targeting while the N-terminus and the central lipophilic core of oleosins contain sequences critical to targeting and stability of oleosins. This knowledge, combined with the evidence that oleosins do not undergo cleavage post-translationally, suggested that it might be possible to replace amino-acid sequences of oleosins with a desired protein or peptide without causing aberrant targeting. That such modifications are possible was shown by Van Rooijen and Moloney (1995a). In those experiments we demonstrated that a large protein such as β-glucuronidase could be made as an inframe fusion with an *Arabidopsis* oleosin expressed in transgenic *Brassica napus*. The resulting fusion protein was accumulated at levels which were about one-tenth that of the native oleosin complement and was essentially all correctly targeted to oil bodies.

This result is in stark contrast to the use of storage proteins as carriers of the fusion protein. In those cases the most predictable outcome would be mistargeting and protein instability (Hoffman *et al.,* 1988). Perhaps the most surprising finding is that correct targeting is apparently independent of the *size* of the C-terminal tag, which in this case is over three times the size of the oleosin. Through this form of lipophilic targeting, Van Rooijen and Moloney (1995a) were able to show that the fusion protein oleosin-β-glucuronidase could be separated from all other cellular proteins except endogenous oleosins by low speed centrifugation. However, an oleosin-protein fusion may not be appropriate for many uses where the purified recombinant protein is needed. To address this problem Van Rooijen and Moloney (1995a) interposed an endoproteolytic cleavage site between the oleosin domain and the β-glucuronidase domain of the fusion protein. The cleavage site was introduced in such a way that according to the theoretical models of oleosin disposition on oil bodies the site should reside on the *outside* of the oil body and be accessible to the endoprotease. The prediction proved to be correct and oil bodies containing oleosin-GUS on their surface released a protein of the correct molecular weight when treated with the endoprotease thrombin. This experiment illustrates the potential use of such a system for the production, separation and release of the desired protein.

It is noteworthy that the oleosin-GUS fusion protein was enzymatically active on oil bodies, a surprising finding, given the widely held assumption that β-glucuronidase

functions as a homo-tetramer. However, it was shown that oil bodies equipped with such a fusion protein act as immobilization matrices for the conversion of a substrate to product. The 'immobilized' catalytic oil bodies can be recovered by centrifugation or pH shift and transferred to a new vessel, where they catalyze the conversion of a second batch of substrate to product without loss of efficiency. This property of oil bodies might be exploited in the production of immobilized enzymes where immobilization frequently costs more per unit of enzyme than does production of the enzyme itself (Daniels, 1987).

The development of this system for the production of industrially or medically useful proteins and peptides is underway. Parmenter *et al.* (*in press*) have recently demonstrated the flexibility of the system for the production of the leech anticoagulant protein *hirudin* (Dodt *et al.*, 1984). Hirudin is an interesting molecule in this regard as it is a relatively short (65 amino acids), hydrophilic protein with three disulfide bridges and thus requires a specific intramolecular conformation for activity. Parmenter *et al.* (1995) show that such a molecule may be produced as a fusion protein which, when it is subjected to proteolytic cleavage off oil bodies, results in fully-functional hirudin of specific activity comparable with that obtained from transgenic yeast secreting recombinant hirudin into the medium. Given the costs associated with producing a metric tonne of an oil seed such as *Brassica napus* and the apparent stability of the fusion in seeds it should be feasible to produce peptides of this nature in seeds with economics comparing favourably to fermentation systems. As the seeds can be stored for at least three years without loss of recombinant protein, it is not necessary to synchronize production with a growing season.

The major drawback to this system is the use of a cleavage enzyme to release the desired product. Enzymes with the required specificity are rare in nature and are frequently quite expensive. Clearly, it will be important for the widespread use of this system to access relatively inexpensive sources for such cleavage enzymes. Although the production of cheap recombinant forms of blood clotting proteases such as thrombin and Factor Xa are still not established, other proteases such as collagenases from micro-organisms (Takeuchi *et al.*, 1992) or plant virus proteases such as those involved in the processing of the Tobacco Etch Virus polyprotein (Rorrer *et al.*, 1992) might prove suitable in many applications.

Concluding remarks

The potential for exploiting plants as production vehicles for recombinant proteins is beginning to be realised. There has been a long period between the initial development of plant transformation and experiments which demonstrate the usefulness of plants as 'solar fermenters'. The nature of products which might become available through plant biosynthesis is diverse. Furthermore, the possibility for added value to the fundamental activities of plant agriculture is enormous. However, it is clear from the discussion above that success in this area will involve a better understanding of gene regulation and the process of subcellular targeting in plant tissues and organs. Significant progress to date has revolved around these two principle areas of study.

Our increased understanding of the secretion pathway in plants has enabled researchers to produce a variety of recombinant proteins from bacteria, fungi and

mammals. Surprisingly, assembly of fairly complex proteins appears to be quite feasible *in vivo*.

The use of lipophilic partitioning systems also holds out hope that a wide variety of proteins, even those of high molecular weight, might be synthesized and easily separated from other cellular proteins. Obviously, our incomplete knowledge of plant protein synthesis, modification and turnover is an impediment to production of certain proteins. Further work is essential in the area of protein turnover, post-translational modification and also subsequent protein assembly.

In this review, we have focused on some of the key experiments which give rise to optimism that plant 'molecular farming' might prove to be a practical alternative to production through fermentation technology. Despite the most important challenges discussed above, it is not unreasonable to predict that the commercialization of plant-based protein production systems will be well underway by the year 2000 and probably widespread as a routine system soon thereafter.

References

AMPE, C., VAN DAMME, J.,DE CASTRO, L., SAMPAIO, M.J., VAN MONTAGU, M. AND VANDEKERCKHOVE, J. (1986). The amino-acid sequence of the 2S sulfur-rich proteins from seeds of Brazil nut (Bertholletia excelsa H.B.K.). *European Journal of Biochemistry* **159**, 597–604.

BAGGA, S., ADAMS, H., KEMP, J.D. AND SENGUPTA-GOPALAN, C. (1995). Accumulation of 15-kilodalton zein in novel protein bodies in transgenic tobacco. *Plant Physioogy*. **107**, 13–23.

BIRD, R.E., HARDMAN, K.D., JACOBSON, J.W., JOHNSON, S., KAUFMAN, B.M., LEE, S-M., LEE, T., POPE, S.H., RIORDAN, G.S. AND WHITLOW, G. (1988). Single-chain antigen binding proteins. *Science* **242**, 423–426.

CRAWLEY, M.J., HAILS, R.S., REES, M., KOHN, D. AND BUXTON, J. (1993). Ecology of transgenic rapeseed in natural habitats. *Nature* **363**, 620–623.

DANIELS, M.J. (1987). Industrial operation of immobilized enzymes. *Methods in Enzymology* **136**, 371–379.

DODT, J., MULLER, H.P., SEEMULLER, U. AND CHANG, J.-Y. (1984). The complete amino acid sequence of hirudin, a thrombin specific inhibitor. *FEBS Letters* **165**, 180–183.

DÜRING, K., HIPPE, S., KREUZALER, F. AND SCHELL, J. (1990). Synthesis and self-assembly of a functional monoclonal antibody in transgenic *Nicotiana tabacum*. *Plant Molecular Biology* **15**, 281–293.

FALCO, S.C., GUIDA, T., LOCKE, M., MAUVAIS, J., SANDERS, C., WARD, R.T. AND WEBBER, P. (1995). Transgenic canola and soybean seeds with increased lysine. *Bio/Technology* **13**, 577.

FAYE, L. (1988). Structure, biosynthesis, and function of plant glycoprotein glycans. *Current Topics in Plant Biochemistry and Physiology* **7**, 62–82.

FAYE, L., JOHNSON, K.D., STURM, A. AND CHRISPEELS, M.J. (1989). Structure, biosynthesis, and function of asparagine-linked glycans on plant glycoproteins. *Physiologia Plantarum* **75**, 309–314.

FAYE, L., GORMOND, V., FITCHETTE-LAINE, A.-C. AND CHRISPEELS, M.J. (1993). Affinity purification of antibodies specific for Asn-linked glycans containing α1fi2 xylose. *Analytical Biochemistry* **209**, 104–108.

FIREK, S., DRAPER, J., OWEN, M.R.L., GANDECHA, A., COCKBURN, B. AND WHITELAM, G.C. (1993). Secretion of a functional single-chain Fv protein in transgenic tobacco plants and cell suspension cultures. *Plant Molecular Biology* **23**, 861–870.

FRALEY, R.T., ROGERS, S.G. AND HORSCH, R.B. (1986). Genetic transformation in higher plants. *CRC Critical Reviews in Plant Science* **4**, 1–46.

GORDON-KAMM, W.J., SPENCER, T.M., MANGANO, M.L., ADAMS, T.R., DAINES, R.J., START, W.G., O'BRIEN, J.V., CHAMBERS, S.A., JR., ADAMS, W.R., JR., WILLETTS, N.G., RICE,

T.B., MACKEY, C.J., KRUEGER, R.W., KAUSCH, A.P. AND LEMAUX, P.G. (1990). Transformation of maize cells and regeneration of fertile transgenic plants. *Plant Cell* **2**, 603–618.

GOTO, M., AKAI, K., MURAKAMI, A., HASHIMOTO, C., TSUDA, E., UEDA, M., KAWANISHI, G., TAKAHASHI, N., ISHIMOTO, A., CHIBA, H. AND SASAKI, R. (1988). Production of recombinant human erythropoietin in mammalian cells: host-cell dependency of the biological activity of the cloned glycoprotein. *Bio/Technology* **6**, 76–71.

HAQ, T.A., MASON, H.S., CLEMENTS, J.D. AND ARNTZEN, C.J. (1995). Oral immunization with a recombinant bacterial antigen produced in transgenic plants. *Science* **268**, 714–716.

HEIN, M.B., TANG, Y., MCLEOD, D.A., JANDA, K.D. AND HIATT, A. (1991). Evaluation of immunoglobulins from plant cells. *Biotechnology Progress* **7**, 455–461.

HIATT, A., CAFFERKEY, R. AND BOWDISH, K. (1989). Production of antibodies in transgenic plants. *Nature* **342**, 76–78.

HILLS, M.J., WATSON, M.D. AND MURPHY, D.J. (1993). Targeting of oleosins to the oil bodies of oilseed rape (*Brassica napus L.*). *Planta* **189**, 24–29.

HOFFMAN, L.M., DONALDSON, D.D., BOOKLAND, R., RASHKA, K. AND HERMAN, E.M. (1987). Synthesis and protein body deposition of maize 15-KDa zein in transgenic tobacco seeds. *EMBO Journal* **6**, 3213–3221.

HOFFMAN, L.M., DONALDSON, D.D. AND HERMAN, E.M. (1988). A modified storage protein is synthesized, processed, and degraded in the seeds of transgenic plants. *Plant Molecular Biology* **11**, 717–729.

HUANG, A.H.C. (1992). Oil bodies and oleosins in seeds. *Annual Reviews of Plant Physiology and Plant Molecular Biology* **43**, 177–200.

JACKS, T.J., HENSARLING, T.P., NEUCERE, J.N., YATSU, L.Y. AND BARKER, R.H. (1990). Isolation and physicochemical characterization of the half-unit membranes of oilseed lipid bodies. *JAOCS* **67**, 353–361.

KEEGSTRA, K. (1989). Transport and routing of proteins into chloroplasts. *Cell* **65**, 247–253.

KOZIEL, M.G., BELAND, G.L., BOWMAN, C., CAROZZI, N.B., CRENSHAW, R., CROSSLAND, L., DAWSON, J., DESAI, N., HILL, M., KADWELL, S., LAUNIS, K., LEWIS, K., MADDOX, D., MCPHERSON, K., MEGHJI, M.R., MERLIN, E., RHODES, R., WARREN, G.W., WRIGHT, M. AND EVOLA, S.V. (1993). Field performance of elite transgenic maize plants expressing an insecticidal protein derived from Bacillus thuringiensis. *Bio/Technology* **11**, 194–200.

KRANTZ, S.B. (1991). Erythropoietin. *Blood* **77**, 419–434.

KREBBERS, E., HERDIES, L., CLERCQ, A., SEURINCK, J., LEEMANS, J., DAMME, J., SEGURA, M., GHEYSEN, G., MONTAGU, M. AND VANDEKERCKHOVE, J. (1988). Determination of the processing sites of an *Arabidopsis* 2S albumin and characterization of the complete gene family. *Plant Physiology* **87**, 859–866.

LI, X., WU, Y., ZHANG, D-Z., GILLIKIN, J.W., BOSTON, R.S., FRANCESCHI, V.R. AND OKIRA, T.W. (1993). Rice prolamine protein body biogenesis: a BiP-mediated process. *Science* **262**, 1054–1056.

MA, J.K.-C., LEHNER, T., STABILA, P., FUX, C.I. AND HIATT, A. (1994). Assembly of monoclonal antibodies with IgG1 and IgA heavy chain domains in transgenic tobacco plants. *European Journal of Immunology* **24**, 131–138.

MA, J.K.-C., HIATT, A., HEIN, M., VINE, N.D., WANG, F., STABILA, P., VAN DOLLEWEERD, C., MOSTOV, K. AND LEHNER, T. (1995). Generation and assembly of secretory antibodies in plants. *Science* **268**, 716–719.

MATSUMOTO, S., IKURA, K., UEDA, M. AND SASAKI, R. (1995). Characterization of a human glycoprotein (erythropoietin) produced in cultured tobacco cells. *Plant Molecular Biology* **27**, 1163–1172.

MAZUR, B.J. AND FALCO, S.C. (1989). The development of herbicide resistance crops. *Annual Review of Plant Physiology and Plant Molecular Biology* **40**, 441–470.

MURPHY, D.J. (1993). Structure, function and biogenesis of storage lipid bodies and oleosins in plants. *Progress in Lipid Research* **32**, 247–280.

NAWRATH, C., POIRIER, Y. AND SOMERVILLE, C. (1994). Targeting of the polyhydroxybutyrate biosynthetic pathway to the plastids of *Arabidopsis thaliana* results in high levels of

polymer accumulation. *Proceedings of the National Academy of Sciences USA*. **91**, 12760–12764.

OWEN, M., GANDECHA, A., COCKBURN, B. AND WHITELAM, G. (1992). Synthesis of a functional anti-phytochrome single chain Fv protein in transgenic tobacco. *Bio/Technology* **10**, 790–794.

PARMENTER, D.L., BOOTHE, J.G., VAN ROOIJEN, G.J.H., YEUNG, E.C. AND MOLONEY, M.M. (1995). Production of biologically active hirudin in plant seeds using oleosin partitioning technology. *Plant Molecular Biology* **29**, 1167–1180.

PEDERSON, K., AGROS, P., NARAVANA, S.V.L. AND LARKINS, B.A. (1986). Sequence analysis and characterization of a maize gene encoding a high-sulfur zein protein of M_r 15, 000. *Journal of Biological Chemistry* **261**, 6279–6284.

PEN, J., VAN OOYEN, A.J.J., VAN DEN ELZEN, P.J.M., RIETVELD, K. AND HOEKEMA, A. (1992). Direct screening for high-level expression of an introduced alpha-amylase gene. *Plant Molecular Biology* **18**, 1133–1139.

PEN, J., VAN OOYEN, A.J.J., VAN DEN ELZEN, P.J.M., QUAX, W.J. AND HOEKEMA, A. (1993a). Efficient production of active industrial enzymes in plants. *Industrial Crops and Prod.* **1**, 241–250.

PEN, J., VERWOERD, T., PARIDON, P.A., BEUDEKER, R., ELZEN, P.J.M. *et al.* (1993b). Phytase-containing transgenic seeds as a novel feed additive for improved phosphorus utilization. *Bio/Technology* **11**, 811–814.

POIRIER, Y., DENNIS, D.E., KLOMPARENS, K. AND SOMERVILLE, C.R. (1992). Polyhydoxybutyrate, a biodegradable thermoplastic produced in transgenic plants. *Science* **256**, 520–523.

ROBERTS, M.R., HODGE, R., ROSS, J.H.E., SORENSEN, A., MURPHY, D.J., DRAPER, J. ANDSCOTT, R. (1993). Characterization of a new class of oleosins suggests a male gametophyte-specific lipid storage pathway. *Plant Journal* **3**, 629–636.

RORRER, K., PARKS, T.D., SCHEFFLER, B., BEVAN, M. AND DOUGHERTY, W.G. (1992). Autocatalytic activity of the tobacco etch virus NIa proteinase in viral and foreign protein sequences. *Journal of General Virology* **73**, 775–783.

ROSS, J.H.E., SANCHEZ, J., MILLAN, F. AND MURPHY, D.J. (1993). Differential presence of oleosins in oleogenic seed and mesocarp tissues in. *Plant Science* **93**, 203–210.

SENGUPTA-GOPALAN, C., REICHERT, N.C., BURKER, R.F., HALL, T.C. AND KEMP, J.D. (1985). Developmentally regulated expression of the bean β-phaseolin gene in tobacco seeds. *Proceedings of the National Academy of Sciences USA* **82**, 3320–3324.

SIJMONS, P.C., DEKKER, B.M.M., SCHRAMMEIFER, B., VERWOERD, T.C., VAN DEN ELZEN, P.J.M., HOEKEMA, A., HUISMAN, M.J. AND MOLENDIJK, L. (1990). Production of correctly processed human serum albumin in transgenic plants. *Bio/Technology* **8**, 273–221.

SMITH, C.J.S., WATSON, C.F., RAY, J., BIRD, C.R., MORRIS, P.C., SCHUCH, W. AND GRIERSON, D. (1988). Antisense RNA inhibition of polygalacturonase gene expression in transgenic tomatoes. *Nature* **334**, 724–726.

STASKAWICZ, B.J., AUSUBEL, F.M., BAKER, B.J., ELLIS, J.G. AND JONES, J.D.G. (1995). Molecular genetics of plant disease resistance. *Science* **268**, 661–667.

TAKEUCHI, H., SHIBANO, Y., MORIHARA, K., FUKUSHIMA, J., INAMI, S., KEIL, B., GILLES, A-M., KAWAMOTO, S. AND OKUDA, K. (1992). Structural gene and complete amino acid sequence of *Vibrio alginolyticus* collagenase. *Biochemical Journal* **281**, 703–708.

TÖPFER, R., MARTINI, N. AND SCHELL, J. (1995). Modification of Plant Lipid Synthesis. *Science* **268**, 681–686.

VAN ROOIJEN, G.J.H. AND MOLONEY, M.M. (1995a). Plant seed oil-bodies as carriers for foreign proteins. *Bio/Technology* **13**, 72–77.

VAN ROOIJEN, G.J.H. AND MOLONEY, M.M. (1995b). Structural requirements of oleosin domains for subcellular targeting to the oilbody. *Plant Physiology* **109**, 1353–1361.

VANDEKERCKHOVE, J., DAMME, J., LIJSEBETTENS, M., BOTTERMAN, J., BLOCK, M., VANDEWIELE, M., CLERCQ, A., LEEMANS, J., MONTAGU, M. AND KREBBERS, E. (1989). Enkephalins produced in transgenic plants using modified 2S seed storage proteins. *Bio/Technology* **7**, 929–932.

VASIL, V., SRIVASTAVA, V., CASTILLO, A.M., FROMM, M.E. AND VASIL, I.K. (1993). Rapid

production of transgenic wheat plants by direct bombardment of cultured immature embryos. *Bio/Technology* **11**, 1553–1558.

VOELKER, T.A., WORRELL, A.C., ANDERSON, L., BLEIBAUM, J., FAN, C., HAWKINS, D.J., RADKE, S.E. AND DAVIES, H.M. (1992). Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. *Science* **257**, 72–74.

VON SCHAEWEN, A., STURM, A., O'NEILL, J. AND CHRISPEELS, M.J. (1993). Isolation of a mutant *Arabidopsis* plant that lacks N-acetyl glucosaminyl transferase I and is unable to synthesize golgi-modified complex N-linked glycans. *Plant Physiology* **102**, 1109–1118.

YATSU, I.Y. AND JACKS, T.J. (1972). Spherosome membranes: Half-unit membranes. *Plant Physiol.* **49**, 937–943.

ZOUBENKO, O.V., ALLISON, L.A., SVAB, Z. AND MALIGA, P. (1994). Efficient targeting of foreign genes into the tobacco plastid genome. *Nucleic Acids Research* **22**, 3819–3824.